DOCUMENT RESUME

ED 170 378 TM 009 005

AUTHOR Alkin, Marvin C.
TITLE Using Naturalistic Research for the Study of Evaluation Utilization.
PUB DATE 11 Apr 79
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Decision Making; Educational Assessment; *Evaluation; *Evaluation Methods; Evaluators; *Information Utilization; Interpersonal Relationship; *Research Methodology; Research Utilization
IDENTIFIERS *Naturalistic Research

ABSTRACT

Suggestions for broadening the utilization of evaluation research findings are discussed. Major program changes, based on evaluation, may not become obvious for several years; program actions may contradict findings and still be based on rational decision making processes; evaluation findings may have far-reaching effects which were not recognized and formally accounted for; personal interaction between evaluator and user may help to shape and redirect the evaluation process. In order to take these factors into account, naturalistic research methods, using techniques such as case studies, field investigations, and participant observations, are recommended. The evaluator's approach is one variable of naturalistic research. The evaluator's influence on the utilization of evaluation information may be determined by: role choice; the extent of which user involvement is encouraged; the amount of attention given to the performance on mandated evaluation tasks; rapport established between the evaluator and users; and the extent to which the evaluator facilitated and stimulated the use of information. Evaluation is defined as a dynamic process; the procedures and outcomes of evaluation studies are influenced by multiple factors as well as interactions between them. Five case studies of evaluation utilization are briefly described. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marvin C. Alkin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Using Naturalistic Research for the Study of Evaluation Utilization*

Marvin C. Alkin
Center for the Study of Evaluation
University of California, Los Angeles

ED170378

There are competing views of what constitutes the utilization of evaluation. And, in fact, the extent to which an evaluation utilization researcher holds one or another of these views affects the kind of research which is appropriately conducted.

One view of utilization looks for direct, immediate impact of an evaluation upon critical decisions made about the evaluated program. For example, an evaluation might be conducted of a special mathematics enrichment program for fourth grade students. If the evaluation showed the enrichment program to have little or no benefit beyond that of the usual math curriculum, then evidence of effective utilization, narrowly defined, might be a decision to terminate the experimental enrichment program or at least to take clear and forceful steps to modify it. There is little problem, in theory at least, in identifying this sort of utilization since its existence is verified by determining whether the findings of the evaluation were acted upon in a clear, rational fashion.

---

*Paper presented at the annual meeting of the American Educational Research Association, April 11, 1979, San Francisco, Calif. The activity which is the subject of this report was supported in whole or in part by the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred. Portions of this paper are excerpted from: M. C. Alkin, R. Daillak, & P. White, Using Evaluations: Does Evaluation Make a Difference?, Library of Social Research #76, Beverly Hills: Sage Publications, 1979.

TM009 005

In practice, however, the findings of an evaluation are not unambiguous; determining what constitutes a rational response is highly problematic; and deciding whether the response was made in reaction to the evaluation or to other forces is equally difficult.

The situation is complicated even more when we consider more long-range and sometimes more subtle effects of evaluation. Returning to the example of our math program, suppose that the enrichment program is expanded to include additional fourth grade students with no official effort to revamp the program content and methods. This certainly does not look like utilization of the evaluation. Yet, there may be other information which puts this into a different light. For example, while the decision makers may have read the report seriously and in good faith, they may also be responding to teachers' and principals' reports that the program had some start-up difficulties but is really beginning to jell and that the teaching approaches employed in the program have been a source of positive morale among the math faculty by getting the old and young teachers together to share ideas and enthusiasm (facts which the evaluation report tends to support).

Or, suppose that after two additional years of continued lackluster academic achievement results the decision is made to adopt a commercially available math instruction program which promises improved achievement and employs teaching approaches now already implemented in the program schools. This decision can, in fact, be traced (in part) to the consistently mediocre showing of the old enrichment program, including the first year's evaluation results, along with the evidence of success of the instructional approaches.

A Broadened Definition of Utilization

Several points are illuminated by this scenario, which tend to broaden the conception of utilization considerably and make the utilization researcher's task more challenging. For example, first, it is not enough to ask in September what the effects of the previous academic year's evaluation have been; as our illustration suggests, it may take two or three or more years before major program changes occur, and assessing the many early inputs to such decisions is necessary if we are to obtain a complete understanding of utilization.

A second important observation is that program actions can seemingly contradict an evaluation, yet decision makers may still be acting rationally and in good faith. In our example, the decision makers gave the evaluation a serious hearing, yet in some of their decisions chose to act contrary to the "clear implications" of the findings. This can happen; evaluations can be utilized, in the broader sense of being "listened to," without being obeyed. It is incumbent upon the evaluation utilization researcher to recognize this; the researcher must not infer nonutilization from superficial observations that the "obvious implications" of an evaluation were not acted upon. Researchers need to become familiar with the decision context in detail.

A third observation is that evaluations have influence beyond the formally stated evaluation concerns. In our illustration, team-teaching, staff morale, and renewal of skills of the senior staff were informal foci of the evolving evaluation. The comments of the evaluator on these matters may have filtered down to influence actions at the teacher level by helping to create an administrative climate which supported and fostered such activities; this expanded conception of utilization directs

our attention to these initially unanticipated impacts of the evalution. The utilization researcher must, in short, be attuned to all the various forms of evaluation "fallout."

Finally, there is the matter of evaluation process or flow. In our illustration, the evaluator and the program administrators adapted to and helped to shape and redirect the evaluation process. The initial evaluation concern with student achievement was supplemented by an increasing interest in the program's effects on teaching staff. The evaluator and decision makers jointly led the evaluation in this direction. Had the evaluator chosen to ignore these "peripheral" concerns in favor of the "bottom line" achievement data, and had the evaluator-decision maker relationship been chilled by the apparent decision to "ignore" the first year's achievement data results, then the entire program history might have been altered. Thus, the various forms of utilization are outcomes of the complex, evolving evaluation process. The researcher who truly wishes to understand the "why?" of utilization cannot treat evaluation as a black box with inputs (characteristics, factors, etc.) and outputs (decision), but must open up the evaluation black box and carefully study the interactions of people and events which produce the multiple consequences of evaluation and which give these consequences meaning.

Research Strategy for Studying Utilization

Proceeding from this alternative conception of utilization, I have outlined some of the considerations which should inform the research: the need to attend to consequences over the long term; sensitivity to the context in which program actions are taken, especially including the other influences upon decision making; exploration of all the manifold

consequences of the evaluation, not simply those relating to the initial, formally stated evaluation concerns; and systematic attention to the evaluation as process, as an unfolding social situation guided by the actors according to their individual and joint understandings of the situation.

The list of important considerations guiding our research efforts could be expanded, but that is not necessary. Simply on the basis of those just described, the choice of appropriate research strategies can be reduced to one class: naturalistic research methods, involving such techniques as case studies, field investigations, participant observations, and the like.

Often researchers try to describe situations in terms of inputs and outputs, independent variables and their consequences on dependent variables, but our knowledge of the processes which link inputs to outputs is seldom very complete. When our predictions of what should occur go awry, we are often at a loss to account for the outcomes and retreat into ad hoc remarks about "complex interactions," "intervening variables," or perhaps just "error variance." Naturalistic research, in contrast, concentrates precisely on the unfolding processes which eventuate in observable outcomes. With such a focus on the "stream of action and interpretation," outcome events less often appear as surprises and more often have identifiable histories and can be seen as the understandable product of a sequence of actions and events. This sensitivity of naturalistic research to social process is precisely what is called for in research on evaluation utilization.

Case Studies of Evaluation Utilization

In studies just completed by my colleagues and me, we performed naturalistic research on evaluation and utilization at five local school sites. Each case study focused on a different ESEA Title I or Title IVc program and described a complete and accurate picture of the evaluation process within the program--looking particularly at the persons who shaped that process, how the evaluation fit into the total operation of the school program, and in what way the evaluation influenced decisions made about the program.

A retrospective interview approach to the case studies was selected. This approach involved interviewing, in depth, the operational staff and the evaluator of an educational program which had been selected for study, and as a supplement to the interviews, reviewing documentary evidence such as program proposals, evaluation reports, and the like. The programs selected for study were all in at least their second year of operation, and because the programs were evaluated annually, each had gone through at least one full evaluation cycle. We emphasize this timing factor; by entering a case study site a number of months after the completion of an annual evaluation, we were in a better position to observe the often neglected longer-term effects of the completed evaluation than we would have been had we appeared on the scene just as the evaluation was coming to a close. The specific methodological procedures employed, including site selection, generalizability and validation procedures, are presented in the full report of this study.

A Framework for Studying Utilization

These case studies were the essential raw materials for constructing a conceptual framework of evaluation utilization--more properly, a

framework of factors affecting utilization. This framework was thoroughly grounded in the detailed data of the case studies, and it attempts to capture the complexity of the real world. Our goal was to develop a framework which fit the phenomena of the five cases, rather than filtering the phenomena to fit some preconceived notions about utilization. The framework consists of general categories of variables which described the evaluation situation and had relevance to utilization. In addition, my colleagues and I began to identify, from our case studies, important properties of each category which depicted more detailed aspects of the category. I will examine one of the categories, "evaluator's approach," and the properties within that category.

Category: Evaluator's Approach

The five case studies suggest that the evaluator's approach--the way the evaluator defines his or her task and goes about the evaluation will influence the utilization of the evaluation information. The evaluators studied all had successes: information produced and utilized, users won over to the idea that evaluation could be meaningful and useful to them, etc. Some were more successful and more influential than others, in part due to fortunate circumstance, but also due to the way they approached the evaluation. By studying the five cases, we can attempt to identify aspects of the evaluator's approach which may influence utilization.

First, it may be important to note some of the aspects of the evaluator's approach which we found to have little (or undetermined) impact on utilization. First, none of the five cases involved the application of a formal evaluation model. Our personal experience with other evaluations suggests that few ESEA Title program evaluations do

employ such models; one can only speculate on the effects that the careful use of such models might have. Second, (and contrary to what the literature might suggest), we found little evidence, in our cases, that research rigor was an important factor affecting utilization.

There were, however, a number of properties within the category of "evaluator's approach" which we did find relevant for utilization. Included in this group of properties which found their basis in the field research were: (1) the evaluator's choice of role; (2) the extent to which evaluators sought user involvement in the evaluation process; (3) the amount of attention given to the performance of mandated evaluation tasks; (4) the rapport between evaluators and important users; and (5) the extent to which evaluators sought to facilitate and stimulate the use of information.

Choice of evaluation role appears to derive from a combination of personal and professional considerations, including experience, style, training, and so forth and manifests itself in two ways. The first consideration relates to the kind of function that the evaluator seeks to fulfill (e.g., curriculum specialist, colleague, facilitator, auditor or monitor, judge, researcher, or combinations thereof). Each of these were found to some extent in our case studies. The second manifestation of the choice of role is the choice of audience. That is, the evaluator must make an implicit or explicit commitment of allegiance, so to speak, to a limited number of audiences. The evaluator, then, may see him/herself as a representative of the "public," a representative of the state, of the program managers in general or the program director personally, or of the local site staff. It is the kind of function and the choice of audience together that constitute the evaluator's overall choice of

role. If I may oversimplify, it could be said that utilization will occur to a greater extent when the evaluator has selected as primary audience the user who most wants information and is likely to use it and where the evaluator adopts a role compatible with the information needs of that user.

Evaluators had different views about the desirability of user involvement--some preferring active user participation, other preferring limited, controlled involvement of users in the evaluation process. Generally, those evaluators who defined their role as one of facilitator or colleague sought to involve users to a greater extent both in terms of involvement in the process and by working with users to widen their understanding of evaluation options. The evaluator-as-judge or the evaluator-as-researcher felt less need for involving users to the same extent. Again, a wide range of extent of user involvement was evidenced in our cases.

Another important dimension of the evaluator's approach has to do with their manner of dealing with mandated evaluation tasks. While the mandated tasks facing the evaluator are many, there is, nonetheless, a surprising amount of discretion in dealing with them. As the cases show, the evaluator may allocate his time and effort so that some of the mandated tasks are accomplished quickly and efficiently, leaving sufficient resources to address high priority evaluation needs of targeted users. For example, it was possible in one of our cases (called Rockland) to evaluate the Title I program as required by the state and still be able to conduct an extensive test of one program component, the "Norton" music program. In a number of other cases as well, evaluators were able to comply with the state reporting requirements, conform to the district's

evaluation policies and, at the same time, devote considerable attention to the concerns of local program personnel.

From the case studies, it appears that many of the aspects of the evaluator's approach which have been described are usually accompanied by the development of a sense of rapport between the evaluator and the important users. The rapport can be either personal or professional, although the case studies indicate that the two are closely related. Personal rapport can most often be seen in evaluator-user contacts that are characterized by their frequency, informality, and flexibility; that is, the evaluator and the user seem to enjoy each other's company and are able to extend that compatibility to their discussions of evaluation matters. Professional rapport is much more task oriented; its principal element is a shared interest in the nature of the program and in the means used to evaluate it. At one case study site, called Clayburne, the rapport between the evaluator and the several principals was predominantly professional in nature. The evaluator's expertise in the subject matter field of the program and his strong personal interest in the program fit perfectly with the users' concerns. The result was a rapport, or affinity on program matters that greatly contributed to the use of evaluation; evaluation came to be seen as an integral part of the principal's decision-making processes.

In our case studies, we found differences in the extent to which evaluators viewed facilitating and stimulating the use of the information as a part of their function. When facilitation or stimulation occurred in our cases, it took the form of the evaluator discussing the findings of an evaluation with the user, helping the user to draw implications and recommendations for action from the data, monitoring the results of

any modifications made on the basis of the evaluation, and so forth. Evaluation did not end when the report was handed to the user. Statistics and the other evaluative data were explained at many points during an evaluation, and evaluators who had established personal as well as professional ties to program decision makers seemed to be in a much stronger position for suggesting uses of their efforts.

Our cases showed a strong link between many of the properties in this category. This can be illustrated by two of the case studies. The trust established by the evaluator at Clayburne, along with his frequent discussions with the principal about the evaluation data, made him a chief source of curricular suggestions on career education approaches. In the Bayview case, the evaluator's ability and desire to suggest useful applications of non-mandated information (e.g., classroom observations of teaching strategies) was utilized by the project managers largely on the strength of the evaluator's previously demonstrated responsiveness to the needs of the staff. Without that personal rapport, it is unlikely that the evaluator would have found an audience for his assessment of potentially sensitive areas.

Interrelationships Among the Categories

In my previous discussion, I have described properties and case study events which exemplify them for one category of our conceptual framework. Nevertheless, to fully explain most events within an evaluation, one must refer to several categories simultaneously. Evaluation is a dynamic process, and the events of an evaluation are the product of multiple influences; this was clearly demonstrated within the case studies.

The dynamic interplay among categories can be seen clearly in the initial evaluation planning that goes on between evaluator and program administrators. Here the evaluator's approach interacts with other categories, including the overall orientation of users. The initial user expectations define a beginning set of evaluation options, but working with the user, we found evaluators who expanded the set of potential evaluation activities by exploring questions or concerns that the user may have had about the program. Considering the ways that evaluation might address these concerns, in turn, expanded the user's expectations of evaluation. In Bayview, program administrators expected very little useful assistance from the evaluation, but the evaluator's involvement of them in the evaluation process and his clear desire to be helpful to the program raised their evaluation expectations and their ultimate use of the evaluation.

Another example of the complex interrelationships between categories is provided by our case studies. We found that frequently utilization takes the form of having gradual influence on administrator perceptions of the evaluated program and that evaluative information is interactive with other data sources in becoming utilized. In the Bayview case, it was only when the second year's data came in and tended to confirm the first year's poor performance in reducing truancy that this data was heeded. Until that point, the first year's disappointing performance was dismissed as anomalous. In yet another case, called Garrison, the positive evaluation report in combination with the local credibility of the evaluator and a politically skilled primary user-principal were together a powerful force in attaining utilization.

Final Comment

What is clear to me from all of this is that evaluation utilization does indeed occur, but its forms as well as the forces which lead to utilization are indeed complex. This complexity in combination with our current inadequate understanding of evaluation and utilization requires a methodological procedure sufficiently sensitive to capture the nuances involved--naturalistic research is currently a most appropriate tool for the study of evaluation utilization.

1MCA:a